DYNAMIC SPECIES CO-OCCURRENCE NETWORKS REQUIRE DYNAMIC BIODIVERSITY SURROGATES

Ayesha I.T. Tulloch[1], Iadine Chadès[2], Yann Dujardin[2], Martin J. Westgate[1], Peter W. Lane[1], David Lindenmayer[1]

[1] Fenner School of Environment and Society, The Australian National University, Canberra ACT 2601 Australia; ayesha.tulloch@anu.edu.au; martin.westgate@anu.edu.au; peter.lane@anu.edu.au; david.lindenmayer@anu.edu.au

[2] CSIRO, Brisbane QLD 4000 Australia; Iadine.Chades@csiro.au; Yann.Dujardin@csiro.au

Keywords: complementarity, community assembly, ecological indicator, integer linear programming, optimal monitoring, network theory, Pareto, spatial prioritisation, surrogacy, species co-occurrence network

Type of article: Original Research Paper, Ecography

Word count (abstract only): 300

Word count (including references): 8449

Number of references: 57

Number of figures: 4

Number of tables: 2 (word count 144)

Author for correspondence: Ayesha I.T. Tulloch

Building 141, Fenner School of Environment and Society, The Australian National University, Canberra ACT 2601 Australia

Telephone: +61 2 6125 4108

Fax: +61 2 6125 0746

Email: ayesha.tulloch@anu.edu.au

Abstract

In conservation it is inevitable that surrogates be selected to represent the occurrence of hard-to-find species and find priority locations for management. However, species co-occurrence can vary over time. Here we demonstrate how temporal dynamics in species co-occurrence influence the ability of managers to choose the best surrogate species. We develop an efficient optimisation formulation that selects the optimal set of complementary surrogate species from any co-occurrence network. We apply it to two Australian datasets on successional bird responses to disturbances of revegetation and fire. We discover that a surprisingly small number of species are required to represent the majority of species co-occurrences at any one time. Because co-occurrence patterns are temporally dynamic, the optimal set of surrogates, and the number of surrogates required to achieve a desired surrogacy power, depend on sampling effort and the successional state of a system. Overlap in optimal sets of surrogates for representing 70% of co-occurring species ranges from zero to 57% depending on when the surrogacy decision is made. Surrogate sets representing early successional communities over-estimate the power of surrogacy decisions at later times. Our results show that in dynamic systems, optimal surrogates might be selected in different ways: (1) use short-term monitoring to choose a larger number of static less-informative surrogates; (2) use long-term monitoring to choose a smaller number of static high-power surrogates that may poorly represent early successional co-occurrence; (3) develop adaptive surrogate selection frameworks with high short-term and long-term surrogacy power that update surrogate sets and capture temporal dynamics in species co-occurrence. Our results suggest vigilance is needed when selecting surrogates for other co-occurring species in dynamic landscapes, as selected surrogates from one time may have reduced effectiveness at a different time. Ultimately, decisions that fail to acknowledge dynamic species co-occurrence will lead to uninformative or redundant surrogates.

Introduction

Knowing about when and how species co-occur is a fundamental concept in ecology and conservation (Bascompte 2010, Hille Ris Lambers et al. 2012). Co-occurrence analysis is the study of positive overlaps (e.g. mixed-species feeding aggregations; Krebs 1973) and negative overlaps (e.g. competitive or predatory) in species' environmental requirements and distributions (Gotelli and Ulrich 2010, Neeson and Mandelik 2014). It enables us to test species assembly rules (Fayle et al. 2015, Gotelli et al. 2010), explore food web interactions (Berlow et al. 2009, Memmott 1999), design efficient and representative reserve systems (Moilanen et al. 2009), and optimise the way in which survey funds are spent (MacKenzie and Royle 2005). The influence of negative species interactions on segregating communities has dominated much of the co-occurrence literature (Fayle et al. 2015, Ulrich and Gotelli 2010). Knowledge of positive co-occurrences between species is equally important for prioritising surrogate selection (Morales-Castilla et al. 2015, Tulloch et al. 2013), an approach that is increasingly relied upon by governments and non-government conservation agencies as the funds for ecosystem management and evaluation fail to match budget requirements. Until recently, the large body of literature on species co-occurrence made one major assumption – that co-occurrence networks are static (Poisot et al. 2015). However, both positive and negative relationships between species are rarely set in stone (Araújo et al. 2011, Bascompte 2010). While the strength and direction of species co-occurrence has been shown to vary depending on the time of year, resource availability, extent of niche overlaps, metabolic constraints, evolutionary history, and vulnerability to environmental change (Araújo et al. 2011, Berlow et al. 2009, Burkle et al. 2013), little attention has been paid to how this variability might influence the selection of surrogates for biodiversity conservation.

The primary role of a biotic surrogate is to represent the occurrence of another species (Lindenmayer et al. 2015). The rationale behind selecting a surrogate is that if a species regularly and predictably co-occurs with other species that are hard to detect or expensive to monitor or manage, it is more efficient to focus on that one (Tulloch et al. 2011). By selecting surrogates whose ecological requirements best encompass those of other species, we might infer the occurrence of undetected species and therefore community

composition (Possingham et al. 2007). Information on surrogate distribution can then be used to allocate funds to the locations in the landscape in which both surrogates and co-occurring undetected species might be most efficiently managed or conserved (Margules et al. 2002). Selecting poor surrogates can lead to suboptimal and ineffective management decisions (Tulloch et al. 2013), if the locations managed for selected surrogate species are not representative of the ecological requirements of the remainder of the community (Lentini and Wintle 2015).

Ecological dynamism (i.e. spatio-temporal variation in species co-occurrence and assemblage composition) presents a challenge for finding optimal surrogates to inform conservation decisions. For example, managing threats that predominantly affect species during winter (e.g. seasonal cattle grazing; Siriwardena et al. 2007), might require knowledge only of species co-occurrence at that time of the year. Possibly more important is the fact that, as time passes in dynamic systems, the chance of surrogates accurately representing species co-occurrence relationships at a previous time diminishes (Burkle et al. 2013). It is unsurprising, therefore, that many surrogacy questions remain unanswered. For instance, in a dynamic successional landscape, how effective are surrogates selected using short-term information on species co-occurrences (the early successional community) at representing long-term or late-successional relationships in biological communities? How might we best select surrogates to avoid loss of surrogacy information in dynamic landscapes and minimise the risk of choosing redundant surrogates? Answers to these questions are crucial for effectively allocating management effort to the set of complementary sites or surrogate species that best represents target biodiversity in the landscape (Grantham et al. 2009, Moore et al. 2011).

Any investigation of species surrogate effectiveness requires methods to elucidate and describe species co-occurrence, a topic that has received much research attention (MacKenzie et al. 2004, Neeson and Mandelik 2014, Poisot et al. 2015, Schluter 1984, Veech 2013, Waddle et al. 2010). A number of studies attempted to inform surrogate decisions using distance-based approaches that quantified pairwise species co-occurrence and grouped species by similarities in abundance or function (Cushman et al. 2010, Sutcliffe et al. 2012). The results of these studies were rarely encouraging. Traditional studies had no way

of identifying and excluding species connections that appeared due to chance alone; they rarely explored complementary relationships within the community; and by arbitrarily grouping species, many important co-occurrence relationships may have been obscured. Numerous attempts have been made to improve co-occurrence analysis methods by differentiating significant non-random patterns from random co-occurrence (Ulrich and Gotelli 2010, Veech 2013), but recommendations for using these methods to inform surrogacy decisions are scarce (Lentini and Wintle 2015). One of the most recent advances in co-occurrence has been to quantify the contribution of niche overlap (i.e. the degree of similarity in species-environmental relationships) to co-occurrence through individualistic species distribution modelling that incorporates both environmental drivers and species interactions (Royan et al. 2015). Models might then be used to derive more realistic networks of co-occurrence that assist with predicting changes to the community (Araújo et al. 2011, Gotelli et al. 2010). Network theory is often used as a way to quantify and visualise biotic interactions in food or nesting webs, plant–animal mutualisms, and extinction cascades (Bascompte 2010, Cockle and Martin 2014, Memmott 1999), but despite its advantages over individualistic or distance-based analyses, is rarely used to analyse complex patterns of co-occurrence among species (Araújo et al. 2011). Furthermore, few authors have provided guidance about how to use networks describing the geographic pattern of co-occurrence among species in decision making for managing and monitoring large communities with variable co-occurrence relationships.

In this study, we employ an approach to tackle surrogate selection in dynamic landscapes undergoing successional change. Our goal is to use co-occurrence analysis to find a set of surrogate species that overlap with a target proportion of the species in a particular successional community, and to explore whether community change due to succession changes this surrogate set. We build networks of species co-occurrence representing different successional time periods, then we transform these into “surrogacy networks” that reflect the strength and direction of species co-occurrence during a particular subset of successional time (Lane et al. 2014, Tulloch et al. 2013). Using simple metrics from network theory, we investigate the robustness of co-occurrence relationships over time, and explore how temporal changes in surrogacy networks affect the optimal set of surrogates selected for monitoring and managing

biodiversity. Our objective is typical of conservation agencies seeking to identify management site priorities through the use of surrogates: find the set of species that maximises the complementary surrogacy power of a network (defined as all of the species co-occurrences observed during a particular subset of monitored time at a given site). We devise a new optimisation formulation for finding optimal complementary surrogate sets in any system using a species co-occurrence network, which allows us to tackle large networks representing pairwise relationships of shared habitat use.

We quantify bird species co-occurrence networks in two long-term monitoring programs in south-eastern Australia undergoing changing conditions over time, due in the first case to restoration of woodland habitat, and in the second to fire recovery in a heathland vegetation community. We explore three questions using our dynamic surrogate networks: (1) Do different allocations of monitoring effort over time lead to different species co-occurrence networks and therefore different optimal surrogates? (2) How robust is a given set of optimal surrogates (i.e. those identified from a monitoring dataset spanning a particular time period) over time, in terms of its power to inform on an alternative time period? (3) How might a targeted surrogate selection strategy that acknowledges species co-occurrence dynamics and community succession be used to improve the surrogacy power of decisions compared with static surrogate selection?

Material and Methods

Species co-occurrence approach

We use the approach of Lane et al. (2014) to calculate species co-occurrence and derive the surrogacy value of a species, s_{ij}, which represents the amount of information that surrogate species *i* provides on target species *j*. The input is a presence/absence matrix of species (*m*) by surveys (*q*). The final output is an *m*-by-*m* surrogacy matrix of values s_{ij} for each species in the range [0, 1] quantifying the strength of any positive relationship between species *i* and *j*. When $s_{ij} = 0$, the presence of species *j* is not associated

with the presence of species *i*, whereas a value close to 1 means that species *i* is a good surrogate for species *j*. The surrogacy value of a species for itself, s_{ii}, is 1.

To derive the *m*-by-*m* surrogacy matrix, we first calculate the odds ratios between each pair of species using the R package sppairs (R Core Team 2014, Westgate and Lane 2015). Odds ratios provide information on the strength and direction of species co-occurrence relationships, identified as a key knowledge gap in species co-occurrence studies (Bascompte 2010). We use odds ratios r_{ij} defined as the odds of the first species (*i*) occurring when the second (*j*) does not, divided by the odds of the first species occurring overall. A value of 1 means that the presence of species *i* and species *j* are not associated in the set of surveys, while $r_{ij} < 1$ means that the presence of species *i* is associated with the absence of species *j*. An odds ratio $r_{ij} > 1$ means that species *i* is a potential surrogate for species *j*. Note that unlike correlations, these odds ratios are not symmetrical: r_{ij} may be larger or smaller than r_{ji}, depending on the relative frequency of occurrence of species *i* and *j* (Lane et al. 2014). This is a common and important attribute of species interactions and co-occurrence (Araújo et al. 2011). Our second step is to set all negative and insubstantial species co-occurrences (i.e. those with an odds ratio of between 1/3 and 3) to zero (Lane et al. 2014). This is a little different to studies that use a measure of statistical rather than biological significance to exclude small effects (Araújo et al. 2011, Gotelli and Ulrich 2010), and we run sensitivity analyses to explore the effect of setting this odds ratio threshold on co-occurrence outcomes (Appendix 2). Finally, we convert each positive odds ratio to a value between 0.5 and 1 using the formula $s_{ij} = r_{ij} /(1+ r_{ij})$. This allows all values to be standardised and ensures that the optimisation is not dominated by large odds ratios, which may derive from fortuitous co-occurrence of some moderately rare species.

Calculating surrogacy power

Building on Tulloch et al. (2013), we define $B(Z)$, the benefits of monitoring surrogate species Z (i.e. the surrogacy power of Z) given a set of target species T:

176 $$B(Z) = \sum_{j \in T} max_{i \in Z} s_{ij} \quad with \ Z \subseteq S, \qquad (1)$$

177 where *T* is the fixed set of *m* target species we want to learn from, and *Z* is a set of candidate surrogate
178 species from which we measure the surrogacy power of monitoring. *Z* is a subset of *S,* the fixed set of all
179 the potential surrogate species. In this equation, we account for the surrogate species value s_{ij} that will
180 contribute the most to knowledge about target species (*max* operator), and sum the values across all the
181 target species *j*. For the purposes of this study, we set all target species *T* to be equal to the set of all
182 monitored species *S* (but this could be adjusted if only rare or conservation-dependent species are of
183 target interest, see Tulloch et al. 2013).

184

185 **Finding optimal surrogates to inform decisions**

186 We set an objective of finding the species that are the best complementary surrogates for all other
187 species in the landscape, given a particular budget for any set *Z* of surrogates that could be selected.
188 Complementarity is included to ensure that redundant surrogates (i.e. those providing information similar
189 to another chosen surrogate) are not selected in the best sets.

190 To address our objective, we use the set of all species as target species (set *T*), and all monitored bird
191 species as potential surrogates (set *S*). The best set of species *Z** using surrogacy information is the set
192 that maximises *B*(*Z*), the summed surrogacy value of each species, for a given budget (*cmax*), formally:

193 $$Z^* = argmax_{Z \subseteq S}\{\sum_{j \in T} max_{i \in Z} s_{ij} \text{ such that } cost(Z) \leq cmax\}. \qquad (2)$$

194 For large networks, it is not possible to solve this problem using an exhaustive search (Chadès et al. 2015,
195 Tulloch et al. 2013). We therefore reformulate this problem as an equivalent integer linear programming
196 problem that allows us to address networks of any size (Garfinkel and Nemhauser 1972).

197

198 **Integer linear programming formulation for the complementary set of surrogates**

199 Let *T* be the (fixed) set of *m* target species and *S* be the (fixed) set of *n* surrogate species. Solving the
200 optimisation problem (Eq. 2) is equivalent to solving the following Integer Linear Program:

201 $$max \sum_{i \in S} \sum_{j \in T} s_{ij} x_{ij} \quad (3)$$

202 Subject to the following constraints:

203 $$\forall j \in T, \sum_{i \in S} x_{ij} \leq 1 \quad (4)$$

204 $$\forall i \in S, \forall j \in T, x_{ij} \leq y_i \quad (5)$$

205 $$\forall i \in S, \sum_{i \in S} y_i \, cost(i) \leq cmax \quad (6)$$

206 $$y_i \in \{0,1\}, x_{ij} \in \{0,1\} \quad (7)$$

207 Where every variable y_i is such that $y_i = 1$ if surrogate species *i* is selected and $y_i = 0$ otherwise, and where
208 every variable x_{ij} is such that $x_{ij} = 1$ if surrogate species *i* provides some surrogacy information about target
209 species *j* and $x_{ij} = 0$ otherwise.

210 Constraint (4) represents the need for only one surrogate species to contribute to target species *j*. This
211 constraint differentiates the problem from a standard complementary reserve design problem and is
212 important because it ensures that high surrogacy for a given target species is not the sum of many species
213 with low surrogacy. Constraint (5) forces contributions of surrogate species *i* to every target species *j* to
214 be null if species *i* is not selected. Constraint (6) represents the budget constraint (i.e. the sum of the cost
215 of each selected surrogate species must be less than a *cmax* value). Finally, constraint (7) forces y_i and x_{ij}
216 to take only integer values. Because we are in a maximisation problem, at optimum, the surrogacy
217 information about any given target species *j* will be systematically maximal over all the possible surrogacy
218 information.

219 This linear programming formulation is important because it is an efficient formulation of the
220 optimisation problem seeking to find complementary sets of species (Tulloch et al. 2013) or actions

(Chadès et al. 2015) for monitoring, management, or spatial prioritisation (Moilanen et al. 2009). We provide the CPLEX code of our formulation in the Supplementary Material.

Without loss of generality, we set equal costs for all species, as they are all monitored using the same method. We use equation (3) to calculate the surrogacy power, representing the complementarity sum of the surrogacy information gained across the whole species co-occurrence network given a selected set of surrogates. The maximum value for surrogacy power is equal to the total number of candidate species *m* in each network. We run analyses to see how increasing the budget constraint (here the number of species surrogates selected) changes the composition and the surrogacy power of the selected set. We expect that more species selected will increase the surrogacy power, but that the slope of this relationship likely depends on the data used to derive species co-occurrences. We also expect diminishing returns in surrogacy power with increasing numbers of surrogates selected, due to redundancy (many species provide the same amount of information) and unique individuals (many species provide information only on themselves, and adding unique individuals to the selected set will increase surrogacy power by only a very small amount, 1/*m*).

Case study datasets

We investigate species co-occurrence in two ecosystems in south-eastern Australia with different histories of land management and different bird species assemblages, monitored using repeated 5-minute point count surveys. We record all birds seen or heard within a 50-m radius of permanent plots along a transect at each site between dawn and mid-morning, and repeat the survey on a second day at a different time using a different observer to control for observer heterogeneity and 'day' effects (Lindenmayer et al. 2009). Our survey protocol follows standards that are widely reported in the ecological literature and corrects for false negative detection errors (Tyre et al. 2003). We pool surveys at each plot within a site to result in one set of species detections per site per season.

1. **Southwest Slopes**

The Southwest Slopes is a region of Australian temperate woodlands in southern New South Wales that has been heavily modified due to clearing for agriculture (Lindenmayer et al. 2010). Revegetation through either new plantings on cleared land or as enhancement plantings of existing remnants has occurred on 28 farms across a broad band over 6800km^2 long since the 1990s, with the objective of restoring endangered Box Gum Grassy Woodland communities. Revegetation has resulted in increased woody vegetation cover and changes to key hollow and food resources (Ikin et al. 2014). We are interested in finding the best surrogates for all bird species responding to revegetation over time. We use an extensive longitudinal dataset gathered over 11 years from 2002 to 2013 from repeated surveys of birds on 65 patches (sites) of revegetated woodland (plantings; 708 surveys), in which 150 bird species have been detected (see Supplementary Material for details). Each site is surveyed 6 times per year within a 4-day period. Surveys are conducted in spring, and additional winter surveys conducted during five of the 11 years.

2. Booderee National Park Heathland, Jervis Bay

Booderee National Park (NP) is a 75km^2 IUCN Category I reserve located on the south-east coast of Australia (~35°10'S, 15°40'E). It is co-managed by Parks Australia (a section of the Australian Federal Government's Department of the Environment) and the Wreck Bay Aboriginal Community. The area has a temperate climate with vegetation types ranging from dry heathland to woodland to rainforest, and there is a well-documented fire history, with wildfires burning the Park on average once every 15-20 years as well as controlled burns for biodiversity management (Lindenmayer et al. 2014). A large wildfire burnt 52 % of the Park in 2003, reducing vegetation cover and changing the composition of the plant community (due to the dependence of many Australian heathland plants on fire for flowering and reproduction; Keith et al. 2014). We are interested in finding the best surrogates for all bird species responding to fire and its effects on vegetation. We select the heathland for this study as it is regularly burnt and is the stronghold of the nationally Endangered Eastern Bristlebird *Dasyornis brachypterus*. We use a longitudinal dataset derived from monitoring 26 heathland sites annually over 11 years from 2003 to 2014

(excluding 2008). Each site is surveyed 4 times per year within a 4-day period, detecting 90 bird species over the course of the surveys (260 surveys; see Supplementary Material for details).

Monitoring scenarios

We apply our two case studies to explore whether alternative allocations of monitoring effort over time in a successional landscape lead to different species co-occurrence networks, and test the robustness of a static set of optimal surrogates (i.e. those identified from a monitoring dataset spanning a particular time period) to reducing or increasing the temporal representativeness of data. To do this, we construct six monitoring scenarios that subset the full survey dataset in ways that reflect different decisions about when to select surrogates during a monitoring program. Our scenarios are chosen to represent co-occurrence dynamics resulting from community succession, season (breeding versus non-breeding), or alternative allocations of sampling effort (see Appendix 3 for more details of these scenarios). The scenarios are:

1. Monitor all sites and all years: Use all available survey data (11 years).
2. Increasing extent of short-term monitoring directly after disturbance (community succession dynamics): Use data only from the first (a) 3, (b) 5, (c) 7, or (d) 9 years after disturbance (here fire or planting).
3. Wait for response lag (community succession dynamics): Use data only from the second half of survey years (survey 6-11 years after disturbance).
4. Survey only in one season (seasonal dynamics): Use data only from (a) spring or (b) winter (only for Southwest Slopes dataset, as Booderee NP is monitored only in spring).
5. Reduce frequency (sampling effort): Survey only once every two years.
6. Reduce temporal cover randomly (sampling effort): Randomly select half of all surveys (Southwest Slopes: 354 surveys; Booderee NP heathland: 130 surveys). This simulates an untargeted survey dataset that might be collected by volunteers in an atlassing or citizen science program.

298 Each dataset is first checked for completeness of sampling by fitting the Chao 2 estimator to species-
299 accumulation data (for details see Supplementary Material). Species accumulation curves for each
300 reduced-data scenario indicate that adequate sampling has been completed to detect at least 85% of all
301 species likely to be present in each case study, although 3-year monitoring (scenario 2a) and random
302 monitoring (scenario 6) have high uncertainty in estimates. We use Pearson's product-moment
303 correlation tests to explore the correlation between each surrogacy matrix (traditional distance-matrix
304 approaches are not suited here as the values in the upper triangle of the matrix are not necessarily the
305 same as the lower triangle). To understand whether the fundamental structure of the species co-
306 occurrence networks changes between time periods, we apply simple metrics quantifying the 'degree'
307 (number of positive co-occurrence relationships) and 'strength' (average s_{ij}) of co-occurrences for each
308 species derived from network theory (Araújo et al. 2011, Barrat et al. 2004, Bascompte et al. 2006). We
309 report these metrics for each species for the first 5 years of monitoring (early succession; Scenario 2b)
310 compared with the last 6 years of monitoring (late succession; Scenario 3).

311 To find optimal surrogates for each scenario of dynamics, we rerun the integer linear programming
312 problem with the same objective (Eq. 2) using dataset subsets representing the communities at different
313 times outlined above. The set of optimal surrogate solutions achieving 70% surrogacy power for each
314 scenario are compared using Bray-Curtis dissimilarity indices (Bray and Curtis 1957).

315 For each scenario, we calculate the relative survey effort required to elucidate species co-occurrences in
316 the monitoring scenario, and the value of the associated surrogate decision in terms of surrogacy power.
317 As a limited set of surrogates is often desirable to enable easier communication and stakeholder
318 engagement, we explore return-on-investment trade-offs between the number of surrogates selected by
319 a monitoring approach to achieve a given desired surrogacy power across the whole network, in relation
320 to effort expended in gathering the data (see Appendix 3 of Supplementary Material for calculations). We
321 use the trade-off curves to identify a set of Pareto efficient solutions for informing surrogacy decisions. A
322 surrogate set is Pareto efficient if it is impossible to find another surrogate set that performs better
323 according to one objective and as good or better according to the other objective.

Because the solution to each of the six scenarios of surrogacy selection represents a static representation of species co-occurrences that may be changing into the future, we also investigate the effect of accounting for co-occurrence dynamics when selecting surrogates for management. We compare the surrogacy power of four surrogate selection strategies with objectives of representing 70% of species co-occurrences at the time of surrogate selection: (i) Static early decision: 1-5 years after disturbance (Scenario 2b), (ii) Static late decision: 6-11 years after disturbance (Scenario 3), (iii) Static pooled decision: using all data from 1-11 years after disturbance (Scenario 1), and (iv) Updated decision: aggregating the optimal set selected from 1-5 years post-disturbance with the optimal set from 6-11 years post-disturbance (a selection that accounts for dynamics).

Results

Species co-occurrence across time

Species co-occurrence networks vary depending on the temporal cover of the dataset used (Fig 1, Appendix 1). Capturing only the first 1-5 years (early succession; scenario 2b), or waiting until 6-11 years post-disturbance (late succession; scenario 3; Fig 1), or including data only from one season (Fig S2), result in different representations of species co-occurrence. In the Southwest Slopes, early successional species co-occurrences are less similar to co-occurrences across all years (i.e. scenario 1; Pearson's product-moment correlation coefficient = 0.62) than late successional co-occurrences (Pearson's correlation coefficient = 0.82), with even lower correlation between early and late successional co-occurrence (Table 1, Fig 1a, b). In the Booderee NP heathland, early successional species co-occurrences are more similar to the co-occurrences across all years than late successional co-occurrences (Table 2).

Species co-occurrence also depends on the sampling approach. For the Southwest Slopes, using only a random half of the surveys to inform surrogate decisions results in lower correlation between species surrogacy matrices compared with all other scenarios of reductions in temporal cover (Table 1). This contrasts with Booderee NP co-occurrences that are least similar to the all-data matrix when data are

subset using either short-term monitoring (only the first three years; Pearson's correlation coefficient = 0.70) or sporadic monitoring (once every two years; Pearson's correlation coefficient = 0.71). In comparison, random selection of surveys in Booderee NP is better correlated with the all-data scenario (Pearson's correlation coefficient = 0.76), more poorly correlated with the first half of the monitoring period (Pearson's correlation coefficient = 0.66), and most correlated (compared with other scenarios) with the second half of monitoring (Pearson's correlation coefficient = 0.73; Table 2).

Network metrics reveal that the poor correlations between the first and second time period are driven by a fundamental change in the structure of the co-occurrence networks (Supplementary Material). Between the first 5 years and the last 6 years of monitoring in the Southwest Slopes, 50% of species (75) increase in degree of positive co-occurrence and 42 (28%) decrease (i.e. have fewer positive co-occurrence relationships). In contrast, in the Booderee NP heathland, just over 50% of species (48) decrease in degree of positive co-occurrence and 26 (29%) increase (i.e. have more positive co-occurrence relationships). Similar proportions of changes are discovered for co-occurrence strength in each case study, with positive correlations between the change in degree and the change in co-occurrence strength (Pearson's correlation coefficients 0.53 and 0.47, $P<0.01$, for Southwest Slopes and Booderee NP respectively; Supplementary Material). There is a general trend in both case studies for species with higher degree to decline in the number of positive co-occurrence relationships, and species with lower degree to increase (linear regressions; Southwest Slopes: $y = -0.27x + 13.28$, $R^2 = 0.21$; Booderee NP: $y = -0.23x + 1.57$, $R^2 = 0.20$). This trend is even stronger for the change in species strength over time (linear regressions; Southwest Slopes: $y = -0.86x + 0.55$, $R^2 = 0.40$; Booderee NP: $y = -0.23x + 1.57$, $R^2 = 0.45$).

Optimal surrogates across time

In both case studies, waiting to assess surrogacy until 11 years of co-occurrence data are collected (scenario 1, encompassing both early- and late-successional communities) means that 51% of the surrogacy power to capture all species co-occurrences across time could be achieved by selecting a single

374 species surrogate (Fig 2). In this scenario of pooling all the monitoring data across time, selecting just six
375 surrogate species (4% of the 150 Southwest Slopes species, and 7% of the 90 heathland species) achieves
376 80% of the surrogacy power of the whole community across time.

377 Because the power of each species as a surrogate changes depending on the time within the succession
378 and the corresponding survey dataset used to calculate it (see Appendix 1), the composition of the best
379 set of surrogates also changes over time (Fig 2). The proportion of overlap between the top surrogate
380 species selected to represent 70% of the bird co-occurrence from each scenario is low, ranging from zero
381 to 57% in the Southwest Slopes depending on the length of time over which birds are monitored (average
382 dissimilarity 87%; Fig 2a,b), and from zero to 57% in the Booderee NP heathland recovering after fire
383 (average dissimilarity 80%; Fig 2c,d; see also Supplementary Material Tables S3 and S4).

384 Surrogates chosen by the all-data scenario consistently outperform surrogates chosen from all scenarios
385 with reduced temporal coverage of data, if the objective is to find the best set of surrogates for all species
386 across all monitoring time (Fig S7). Relative to the surrogates selected by the all-data scenario, the
387 scenarios that require the most surrogates to achieve equal surrogacy power to the best set across all
388 time are random monitoring and monitoring only the first three years (the short-term monitoring
389 scenario; Fig 2 and S7).

390 Trading off the number of selected surrogate species (i.e. the surrogate budget) against the effort and
391 time expended on collecting bird survey data results in a Pareto-optimal frontier whose shape depends on
392 the case study and the desired surrogacy power of the selected set (Fig 3). In general, increasing survey
393 effort leads to selecting a smaller set of surrogates that on average provide higher surrogacy power.
394 However, high monitoring effort (in terms of the number of surveys collected to inform surrogacy
395 decisions) does not always minimise the number of surrogates selected for a desired surrogacy power (Fig
396 3). Furthermore, increasing the desired level of surrogacy power across the whole network increases the
397 number of surrogates required to achieve that objective. For example, in Booderee NP, despite low
398 overlap in the composition of the best surrogate set for achieving 70% surrogacy (Fig 2c,d), doubling the
399 surrogate set selected from early succession (from 6 to 12 surrogates) results in the ability to achieve a

400 similar surrogacy power (70% of the entire bird network represented) to that of the all-data scenario (Fig
401 S7).

402 Because the best surrogate sets selected in each scenario may represent only co-occurrences during that
403 particular time period, we compare these static approaches to surrogate selection to a scenario that
404 updates surrogates over time. While it is possible to identify an optimal set of surrogates that represent
405 70% of observed co-occurrences using static approaches of either evaluating co-occurrences 1-5 years
406 (early succession), 6-11 years (late succession), or 1-11 years (early- and late-succession) post-
407 disturbance, each of these surrogate sets has poorer surrogacy power at a different time (Fig 4). For
408 example, the optimal surrogate set for achieving 70% surrogacy power in Booderee NP during the first 5
409 years post-disturbance has considerably lower surrogacy power (51%) over the last 5 years (Fig 4b).
410 Because there is no overlap in the early and late-successional optimal surrogate sets in either case study
411 (Fig 2b,d), we find that the best approach for informing surrogate sets with high surrogacy power in both
412 case studies is to use the updating strategy (Fig 4). This achieves higher surrogacy power (representing up
413 to 90% of species co-occurrence in Booderee NP and 88% in the Southwest Slopes) than the static
414 approaches.

415

416 Discussion

417 Species distributions and abundances are rarely static. Changing resource and shelter availability,
418 competitive and facultative interactions, predation rates, and pressures from anthropogenic threats, lead
419 to spatial and temporal variability in population dynamics and distributions (Burkle et al. 2013, Poisot et
420 al. 2015). Understanding how species co-occurrence changes over time is important for learning about
421 where and when to monitor species, and for ensuring that from a conservation perspective, the right
422 places are protected and managed at any given time. This information becomes more important when we
423 have neither the resources nor the time to learn about the whole system, and must instead rely on
424 surrogates that provide a partial picture of the community. Despite a clear need to understand temporal

dynamics in ecological communities (Poisot et al. 2015), few studies have explored how temporal variability in species' co-occurrence affects decisions for selecting optimal surrogates. We present the first investigation of the effect of temporal dynamics in species co-occurrences on choosing optimal species surrogates. Importantly, we find that the optimal set of complementary surrogates changes over time in successional landscapes.

Conservation budgets are limited, so managers might be tempted to inform surrogate decisions using data collected over a short timeframe, with the assumption that the species selected as surrogates for monitoring or managing would remain representative of other species over time. Our results show that this could be a dangerous strategy in a dynamic landscape. This is because the best set of surrogates is a trade-off between how much data one can afford to collect, how much of the temporal variability in the system one wants captured by the surrogates (e.g. what part(s) of the succession we wish to represent), and the desired degree of surrogacy power (Fig 4). We describe these trade-offs for both of our study locations using Pareto frontiers, and identify the number of surrogates required to achieve a desired surrogacy power over a given time period post-disturbance (Fig 3). Short-term early-successional monitoring strategies require many more surrogates to achieve the equivalent surrogacy power of a few high-power surrogates selected from long-term monitoring data (Fig 2 and 3). Aiming for higher surrogacy power to better represent all species in the network comes at a cost of either an increase in survey effort over time, or an increase in the number of surrogates selected. Accepting more surrogates for selection results in an exponential decline in survey effort required to inform this decision in the Southwest Slopes plantings (Fig 3a), and a decline in survey effort in the Booderee NP heathland that changes from non-linear to linear with increasing desired surrogacy power (Fig 3b). Some monitoring strategies (e.g. 'random' and 1-in-2 year sampling) are suboptimal for finding high-power surrogates compared with strategies of similar cost using continuous data (Fig 3). This is an important result for organisations attempting to assess and select surrogates using citizen science data, as it suggests that medium-term monitoring (between 5 and 9 consecutive years post-disturbance in our study) is more effective for selecting surrogates with high power across time than sporadic or short-term sampling.

Because the optimal set of surrogates changes over time, vigilance is required when setting objectives for surrogate selection to ensure that surrogates are fit for purpose. If we want surrogates that represent only species co-occurrences during periods of resource scarcity or disturbance (e.g. in winter or immediately post-fire), we select a different set of species compared with the surrogates chosen to represent co-occurrences in a different season (Fig S2) or after a response lag (Fig 1). In some cases there is no overlap in the composition of optimal surrogate sets from different time periods (Fig 2). For example, surrogate species selected in the Southwest Slopes (Fig 4a) and in Booderee NP (Fig 4b) to represent species occurrences in the last half of the monitoring years do not overlap with species selected during the first five years. The early-successional surrogates are efficient for representing the bird community immediately post-disturbance, but in the case of Booderee NP, 19% less efficient for representing bird community recovery in the second 5 years than surrogates selected specifically from the late-successional dataset. This means that the best set of surrogates are not only dependent on the system (i.e. location and ecosystem type), but also depend on the successional state of the system (Possingham 1997), i.e. the set of unique biotic and abiotic conditions at a particular point during recovery from disturbance (here fire or restoration).

Updating the best surrogate set to acknowledge dynamics in species co-occurrence improves surrogacy power compared with static surrogate selection. For instance, in the Booderee NP heathland, the surrogacy power of the updated surrogate set is 7% higher than the 1-5 year selection, 3% higher than the 6-11 year selection, and 20% higher than a static selection of surrogates using data pooled over time (Fig 4b). We observe step-wise changes in the best surrogate sets that appear to be stabilising 9 years post-fire in Booderee NP (Fig 2c), but show less evidence of stabilising in the Southwest Slopes (Fig 2a). This suggests that the dynamics of each system operates on a different timescale. In the Booderee NP heathland, this is most likely because the bird dynamics are successional until reaching a relatively stable state (fire is a regular occurrence). The system in the first 3 to 6 years post-fire provides high nectar and pollen resources due to the flowering of fire-dependent shrub species, leading to the need for pollinator surrogates such as the Little Wattlebird and White-cheeked Honeyeater. After vegetation cover has had

time to recover, surrogates for ground-and tree-dwelling species requiring dense cover appear more frequently (e.g. Eastern Bristlebird and Rainbow Lorikeet; Fig 2c). In contrast, restoration in degraded agricultural landscapes can require many more years to achieve biodiversity outcomes and has been shown to result in novel ecosystems in some cases (Lindenmayer et al. 2008), research that is supported by the constantly evolving set of optimal surrogates in the Southwest Slopes (Fig 2a). Knowing when and how state-dependent processes should be incorporated into decision-making is crucial for effective monitoring and management (Hauser et al. 2006). By demonstrating state-dependency in bird surrogate decisions, this study supports previous calls for surrogates to be developed in an adaptive framework (Lindenmayer and Likens 2009). We do not attempt to explain the mechanisms behind these dynamics, and we highlight the need to study the dynamics of co-occurrence network change for other taxonomic groups and in other parts of the world, including elucidating the processes that drive state-dependent changes in co-occurrence networks. In this study, we have identified the surrogates for successional communities monitored for >10 years after a single disturbance event. Ideally, multiple disturbance events should be studied to validate surrogate composition across multiple disruptions. Only with this knowledge might we begin to predict future dynamics of species co-occurrence and community assembly.

Our results support previous findings that a relatively small set of complementary species can provide information about most of the species co-occurrence network at a point in time (Tulloch et al. 2013). We find quickly diminishing returns for surrogate sets of more than six species (which provide information on more than 80% of the network when all data are used in prioritisation), with many species providing occurrence information only on themselves. Complementarity is a concept that has received much attention in ecological decision-making (Justus and Sarkar 2002, Moilanen 2008, Moilanen et al. 2009). Finding the best complementary sets is a challenging optimisation problem due to the exponentially increasing decision space for large datasets of co-occurrences (Chadès et al. 2015). We test our new formulation using two case studies of long-term bird monitoring in Australian ecosystems that allow a range of scenarios of temporal representativeness to be explored. We focus on the surrogacy value of a species (i.e. a network node) for the occurrence of every other species within a successional community.

503 Unlike other studies, we do not attempt to infer biotic interactions from our co-occurrence matrices
504 (Morales-Castilla et al. 2015). Recently, a probabilistic framework was proposed for building interaction
505 networks from co-occurrence (Cazelles et al. 2015), which relies on gathering data on species interactions
506 as well as co-occurrence to build conditional probabilities of biotic interactions (data that were not
507 available in this study). Despite its large data requirements, linking biotic interactions to networks of
508 species co-occurrence remains an important area of future work. However, because of the way our
509 problem is formulated and the way that odds ratios are calculated, we are able to exclude biotic
510 interactions such as competition from our networks, and our maximization algorithm attempts to find
511 values where the interaction strength for species A with B is high (thereby potentially focusing mainly on
512 mutualistic or commensal relationships (Morales-Castilla et al. 2015)). Our formulation is generalizable
513 across species and systems, and can be used to deal with small or large networks in other contexts, for
514 example in finding the optimal set of sites to manage across a network of patches, selecting key nodes to
515 manage within food webs, and discovering keystone species (Berlow et al. 2009, Chadès et al. 2015).

516 Several assumptions are made in this study for the purposes of clarifying the role of data availability in
517 surrogacy decisions and demonstrating the method. First, we assume equal monitoring costs across
518 species – a necessary simplification to allow us to compare the value of different kinds of surrogacy
519 information. Our method allows for the true costs of gathering data using different protocols or
520 equipment to be used as inputs (Gardner et al. 2008). We do not account for differing species
521 detectability as our monitoring protocols are set up to deal with detectability issues through standardised
522 repeated sampling that minimises the risk of imperfect detection issues (Lindenmayer et al. 2009). Studies
523 applying our methods that wish to account for detectability can explore the sensitivity of optimal
524 surrogate sets by incorporating species detectability as a weighting factor into our benefit function
525 (equation 1) (Tulloch et al. 2011, Tulloch et al. 2013). Our odds ratio measure of species co-occurrence
526 based on binary presence-absence matrices of detections is one of many approaches to derive species co-
527 occurrence; some of which rely on knowledge that more abundant species are more likely to co-occur
528 (Berlow et al. 2009, Ulrich and Gotelli 2010), others that incorporate models of trait-based and biotic

529 modifiers of co-occurrence to find biologically meaningful relationships (Poisot et al. 2015), and yet others
530 that use statistical tests to distinguish non-random from random associations (Gotelli and Ulrich 2010).
531 We should note that, like many other co-occurrence measures, our use of an odds ratio threshold to
532 indicate 'strong' positive co-occurrences is vulnerable to Type I errors – assuming a meaningful
533 association is occurring when it is not (Gotelli and Ulrich 2010). Because our optimisation formulation
534 accommodates any alternative measure that provides a relationship between surrogates and targets, we
535 explore the impact of Type I errors on the outcomes of our surrogacy optimisation in sensitivity analyses
536 (Appendix 2.3). These analyses show that the composition of the optimal surrogate set is dependent on
537 how co-occurrence is measured, and associated with this, the willingness of the decision-maker to accept
538 Type I errors brought about by the inclusion of random co-occurrences in the input matrix. Importantly,
539 however, our findings of change in surrogate sets over time are robust to the approach used to derive co-
540 occurrence, strengthening our conclusions that the optimal surrogate set depends on the successional
541 state of the landscape.

542 Our results demonstrate that surrogacy decisions are state-dependent and that biodiversity surrogate
543 information should be updated over time. Careful consideration of time-frames, goals and desired
544 conservation outcomes is needed when selecting surrogates in dynamic landscapes. Reducing the amount
545 of data available to inform surrogacy decisions often decreases the efficiency of the selected surrogacy
546 set. Despite this, decisions must be made, typically without long-term monitoring datasets (Martin et al.
547 2012). Our study suggests that at least in the successional systems we studied, short- to medium-term
548 annual monitoring of more species provides the same information on species co-occurrence as longer-
549 term monitoring with fewer surrogates. Surrogate choices that fail to acknowledge dynamics in species
550 co-occurrence could lead to changes in species assemblages being missed due to surrogates being chosen
551 that are not representative of the whole system. Our results suggest that managers have three choices for
552 selecting surrogates in dynamic landscapes. These are: (1) use short-term monitoring immediately post-
553 disturbance to select a larger number of static less-informative surrogates that are highly representative
554 of early successional co-occurrences but may not be robust over time; (2) use long-term monitoring

555 across early- and late-succession to select a smaller number of static highly informative surrogates that
556 minimise management effort but poorly represent some short-term co-occurrences, or (3) develop
557 adaptive surrogate selection frameworks with high short-term and long-term surrogacy power that
558 update surrogate sets by accounting for successional dynamics in species co-occurrence. Considering
559 temporal dynamics in species co-occurrence will ensure that the best set of complementary surrogates is
560 selected to represent responses in species networks to management.

561

562 Acknowledgements

563 We thank Arthur Le Rhun for assistance with initial programming, Jennifer Pierson, Christopher
564 MacGregor and Ross Cunningham for early discussions, and field staff and volunteers for collecting bird
565 data. This work was supported by the Australian Government's National Environmental Research Program
566 (NERP; AITT), a CSIRO Julius Career Award (IC), a CSIRO OCE Postdoctoral Fellowship (YD) and an
567 Australian Research Council Laureate Fellowship (DL). All bird monitoring was undertaken with animal
568 ethics approval obtained through The Australian National University Animal Experimentation Ethics
569 Committee. The Southwest Slopes study was conducted on privately-owned land and access permission
570 granted by landowners prior to establishing field sites.

References

Araújo, M. B. et al. 2011. Using species co-occurrence networks to assess the impacts of climate change. — Ecography 34: 897-908.

Barrat, A. et al. 2004. The architecture of complex weighted networks. — P Natl Acad Sci USA 101: 3747-3752.

Bascompte, J. 2010. Structure and Dynamics of Ecological Networks. — Science 329: 765-766.

Bascompte, J. et al. 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. — Science 312: 431-433.

Berlow, E. L. et al. 2009. Simple prediction of interaction strengths in complex food webs. — Proc. Natl. Acad. Sci. 106: 187-191.

Bray, J. R. and Curtis, J. T. 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. — Ecol Monogr 27: 325-349.

Burkle, L. A. et al. 2013. Plant-Pollinator Interactions over 120 Years: Loss of Species, Co-Occurrence, and Function. — Science 339: 1611-1615.

Cazelles, K. et al. 2015. A theory for species co-occurrence in interaction networks. — Theor Ecol 1-10.

Chadès, I. et al. 2015. Benefits of integrating complementarity into priority threat management. — Conserv. Biol. 29: 525–536.

Cockle, K. L. and Martin, K. 2014. Temporal dynamics of a commensal network of cavity-nesting vertebrates: increased diversity during an insect outbreak. — Ecology 96: 1093-1104.

Cushman, S. A. et al. 2010. Use of Abundance of One Species as a Surrogate for Abundance of Others. — Conserv. Biol. 24: 830-840.

Fayle, T. M. et al. 2015. Experimentally testing and assessing the predictive power of species assembly rules for tropical canopy ants. — Ecol. Lett. 18: 254-262.

Gardner, T. A. et al. 2008. The cost-effectiveness of biodiversity surveys in tropical forests. — Ecol. Lett. 11: 139-50.

Garfinkel, R. S. and Nemhauser, G. L. 1972. Integer programming. — Wiley New York.

597 Gotelli, N. J. et al. 2010. Macroecological signals of species interactions in the Danish avifauna. — P Natl
598 Acad Sci USA 107: 5030-5035.
599 Gotelli, N. J. and Ulrich, W. 2010. The empirical Bayes approach as a tool to identify non-random species
600 associations. — Oecologia 162: 463-477.
601 Grantham, H. S. et al. 2009. Delaying conservation actions for improved knowledge: how long should we
602 wait? — Ecol. Lett. 12: 293-301.
603 Hauser, C. E. et al. 2006. Should managed populations be monitored every year? — Ecol. Appl. 16: 807-
604 819.
605 Hille Ris Lambers, J. et al. 2012. Rethinking Community Assembly through the Lens of Coexistence Theory.
606 — Annu. Rev. Ecol. Evol. Syst. 43: 227-248.
607 Ikin, K. et al. 2014. Multi-Scale Associations between Vegetation Cover and Woodland Bird Communities
608 across a Large Agricultural Region. — Plos One 9: e97029.
609 Justus, J. and Sarkar, S. 2002. The principle of complementarity in the design of reserve networks to
610 conserve biodiversity: a preliminary history. — J. Biosoc. 27: 421-435.
611 Keith, D. et al. 2014. Heathlands. — In: Lindenmayer, D. et al. (eds), Biodiversity and Environmental
612 Change. CSIRO Publishing.
613 Krebs, J. R. 1973. Social learning and the significance of mixed-species flocks of chickadees (Parus spp.). —
614 Canadian Journal of Zoology 51: 1275-1288.
615 Lane, P. W. et al. 2014. Visualization of species pairwise associations: a case study of surrogacy in bird
616 assemblages. — Ecol Evol 4: 3279-3289.
617 Lentini, P. E. and Wintle, B. A. 2015. Spatial conservation priorities are highly sensitive to choice of
618 biodiversity surrogates and species distribution model type. — Ecography doi:
619 10.1111/ecog.01252.
620 Lindenmayer, D. et al. 2010. Temperate woodland conservation and management. — CSIRO Publishing.
621 Lindenmayer, D. et al. 2014. Booderee National Park. — CSIRO Publishing.
622 Lindenmayer, D. et al. 2015. A new framework for selecting environmental surrogates. — Sci. Total
623 Environ. 538: 1029-1038.

624 Lindenmayer, D. B. et al. 2008. Novel ecosystems resulting from landscape transformation create
625 dilemmas for modern conservation practice. — Conservation Letters 1: 129-135.
626 Lindenmayer, D. B. and Likens, G. E. 2009. Adaptive monitoring: a new paradigm for long-term research
627 and monitoring. — Trends Ecol Evol 24: 482-486.
628 Lindenmayer, D. B. et al. 2009. Do observer differences in bird detection affect inferences from large-scale
629 ecological studies? — Emu 109: 100-106.
630 MacKenzie, D. I. et al. 2004. Investigating species co-occurrence patterns when species are detected
631 imperfectly. — J. Anim. Ecol. 73: 546-555.
632 MacKenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies: general advice and allocating survey
633 effort. — J. Appl. Ecol. 42: 1105-1114.
634 Margules, C. R. et al. 2002. Representing biodiversity: Data and procedures for identifying priority areas
635 for conservation. — J Biosci 27: 309-326.
636 Martin, T. G. et al. 2012. Acting fast helps avoid extinction. — Conservation Letters 5: 274-280.
637 Memmott, J. 1999. The structure of a plant-pollinator food web. — Ecol. Lett. 2: 276-280.
638 Moilanen, A. 2008. Generalized Complementarity and Mapping of the Concepts of Systematic
639 Conservation Planning. — Conserv. Biol. 22: 1655-1658.
640 Moilanen, A. et al. 2009. Spatial Conservation Prioritization: Quantitative Methods and Computational
641 Tools. — Oxford University Press.
642 Moore, J. L. et al. 2011. Estimating detection-effort curves for plants using search experiments. — Ecol.
643 Appl. 21: 601-607.
644 Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. — Trends Ecol Evol 30: 347-356.
645 Neeson, T. M. and Mandelik, Y. 2014. Pairwise measures of species co-occurrence for choosing indicator
646 species and quantifying overlap. — Ecological Indicators 45: 721-727.
647 Poisot, T. et al. 2015. Beyond species: why ecological interaction networks vary through space and time.
648 — Oikos 124: 243-251.

649 Possingham, H. P. 1997. State-dependent decision analysis for conservation biology. — In: Pickett, S. T. A.
650 et al. (eds), The ecological basis of conservation: heterogeneity, ecosystems and biodiversity.
651 Chapman and Hall, pp. 298-304.
652 Possingham, H. P. et al. 2007. How can you conserve species that haven't been found? — J Biogeogr 34:
653 758-759.
654 R Core Team 2014. R: A language and environment for statistical computing, reference index version
655 3.1.1. — R Foundation for Statistical Computing.
656 Royan, A. et al. 2015. Shared environmental responses drive co-occurrence patterns in river bird
657 communities. — Ecography doi: 10.1111/ecog.01703.
658 Schluter, D. 1984. A Variance Test for Detecting Species Associations, with Some Example Applications. —
659 Ecology 65: 998-1005.
660 Siriwardena, G. M. et al. 2007. The effect of supplementary winter seed food on breeding populations of
661 farmland birds: evidence from two large-scale experiments. — J. Appl. Ecol. 44: 920-932.
662 Sutcliffe, P. R. et al. 2012. Biological surrogacy in tropical seabed assemblages fails. — Ecol. Appl. 22:
663 1762-1771.
664 Tulloch, A. et al. 2011. Wise selection of an indicator for monitoring the success of management actions.
665 — Biol. Conserv. 144: 141-154.
666 Tulloch, A. I. T. et al. 2013. Accounting for Complementarity to Maximize Monitoring Power for Species
667 Management. — Conserv. Biol. 27: 988-999.
668 Tyre, A. J. et al. 2003. Improving precision and reducing bias in biological surveys: Estimating false-
669 negative error rates. — Ecol. Appl. 13: 1790-1801.
670 Ulrich, W. and Gotelli, N. J. 2010. Null model analysis of species associations using abundance data. —
671 Ecology 91: 3384-3397.
672 Veech, J. A. 2013. A probabilistic model for analysing species co-occurrence. — Glob. Ecol. Biogeogr. 22:
673 252-260.
674 Waddle, J. H. et al. 2010. A new parameterization for estimating co-occurrence of interacting species. —
675 Ecol. Appl. 20: 1467-1475.

676 Westgate, M. J. and Lane, P. W. 2015. sppairs: Species Pairwise Association Analysis in R. version 0.2,
677 available online: https://github.com/mjwestgate/sppairs.

678

679 # Tables

680 **Table 1.** Results (correlations) of pairwise Pearson's correlation tests for the Southwest Slopes species co-
681 occurrence matrices, representing different scenarios of reductions in the temporal cover of data. For all
682 correlations P<0.01.

Dataset	All data (1–11 yrs)	1–3 yrs	1–5 yrs	1–7 yrs	1–9 yrs	6–11 yrs	Spring	Winter
1–3 yrs	0.41	1						
1–5 yrs	0.62	0.59	1					
1–7 yrs	0.65	0.58	0.54	1				
1–9 yrs	0.67	0.53	0.50	0.91	1			
6–11 yrs	0.82	0.30	0.40	0.46	0.54	1		
Spring	0.85	0.42	0.64	0.59	0.62	0.69	1	
Winter	0.59	0.34	0.48	0.36	0.38	0.60	0.34	1
Random	0.47	0.51	0.40	0.59	0.63	0.46	0.42	0.42

683

Table 2. Results (correlations) of pairwise Pearson's correlation tests for Booderee NP heathland species co-occurrence matrices, representing different scenarios of reductions in the temporal cover of data. For all correlations P<0.01.

Dataset	All data (1–11 yrs)	1–3 yrs	1–5 yrs	1–7 yrs	1–9 yrs	6–11 yrs	1 in 2 yrs
1–3 yrs	0.70	1					
1-5 yrs	0.82	0.84	1				
1–7 yrs	0.87	0.80	0.95	1			
1–9 yrs	0.92	0.75	0.89	0.94	1		
6–11 yrs	0.74	0.47	0.47	0.54	0.64	1	
1 in 2 yrs	0.80	0.75	0.77	0.73	0.77	0.57	1
Random	0.76	0.58	0.66	0.68	0.71	0.73	0.64

Figure Captions

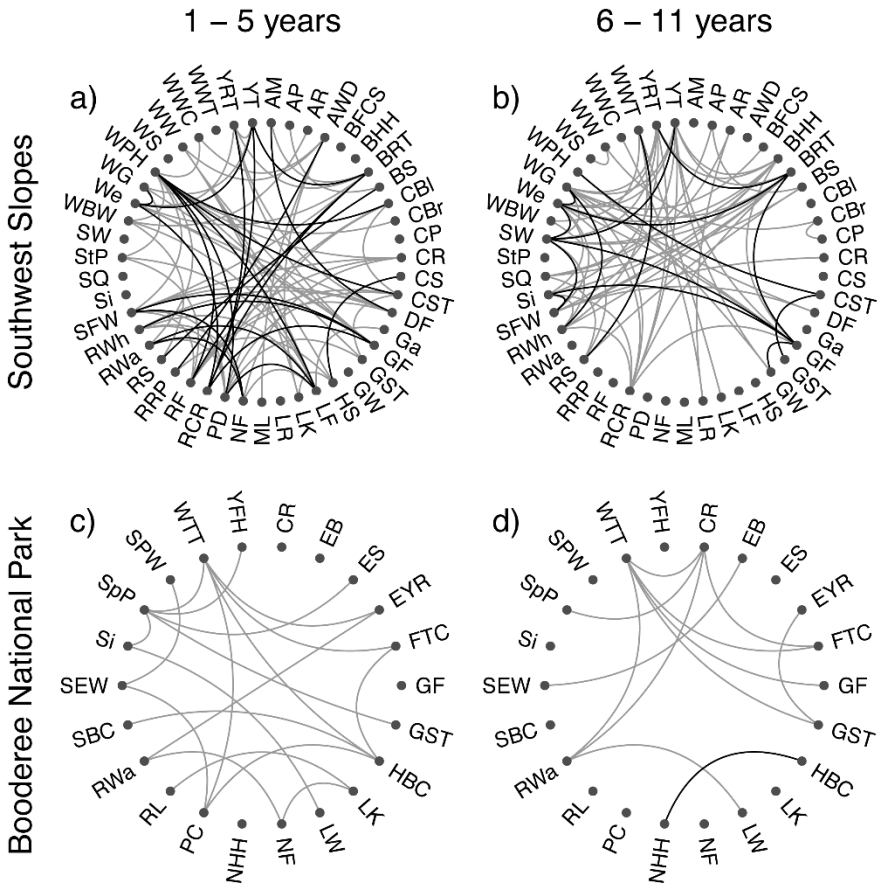


Figure 1. Species co-occurrence networks in the Southwest Slopes plantings derived from (a) early succession (first 6 years) and (b) late succession (next 6 years) after revegetation, and for the Booderee National Park heathland (c) early succession and (d) late succession post-fire. Only showing species detected in >5% of surveys with odds ratio threshold of >3 for strong positive effects (grey lines), with odds ratio of >6 represented by black lines. See Appendix 1 of Supplementary Material for key to labels and matrices and networks representing other scenarios of co-occurrence dynamics.

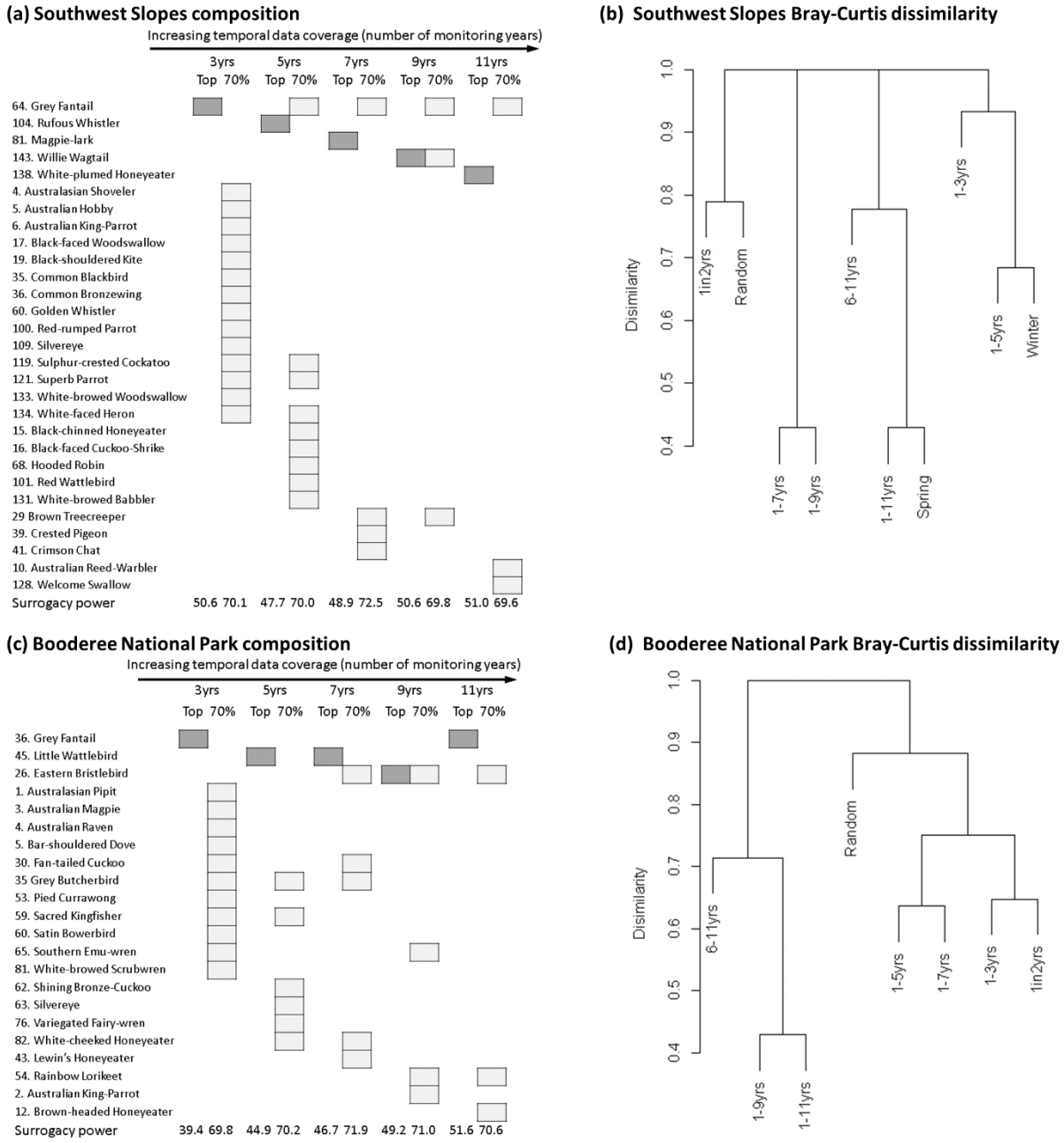
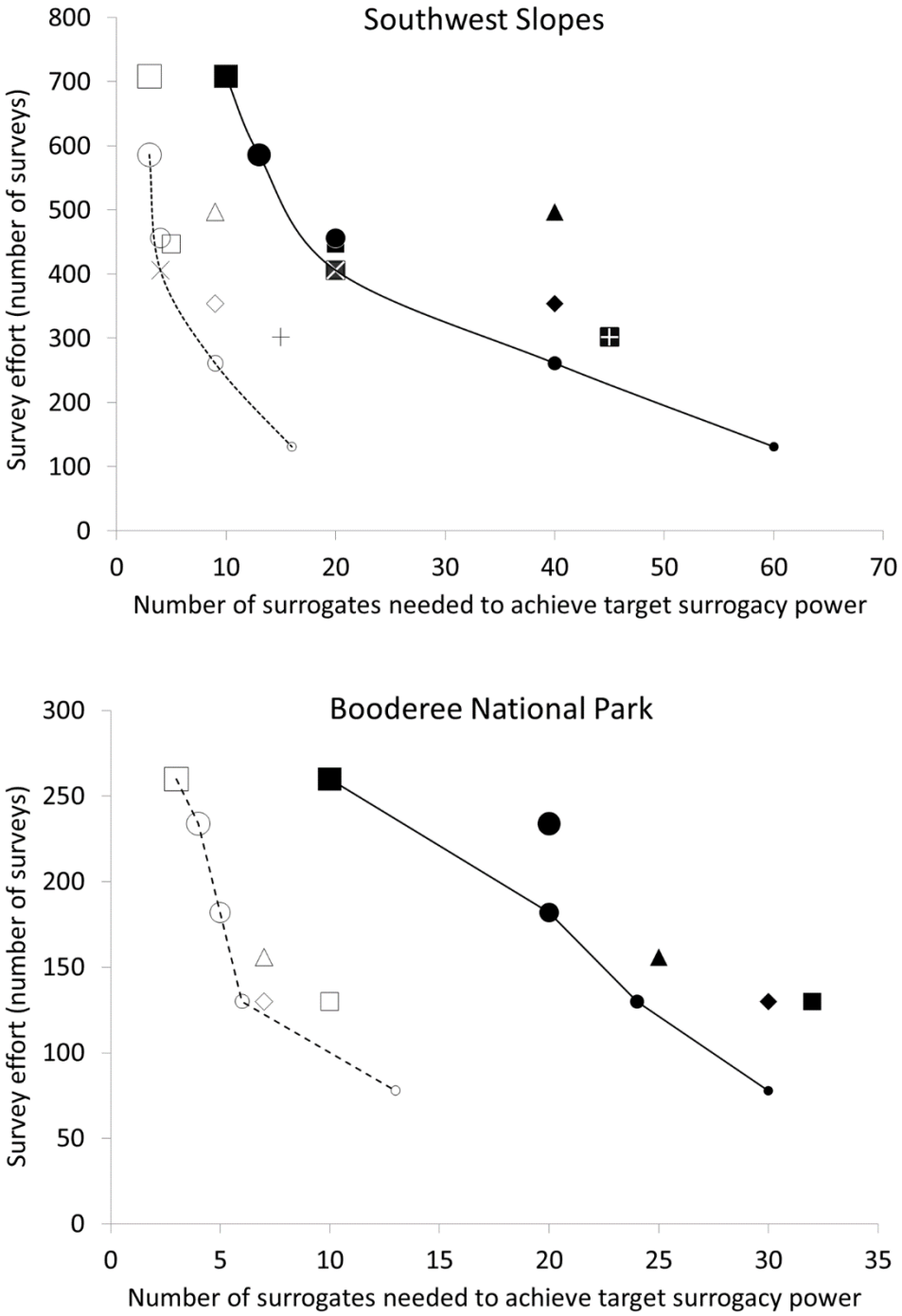
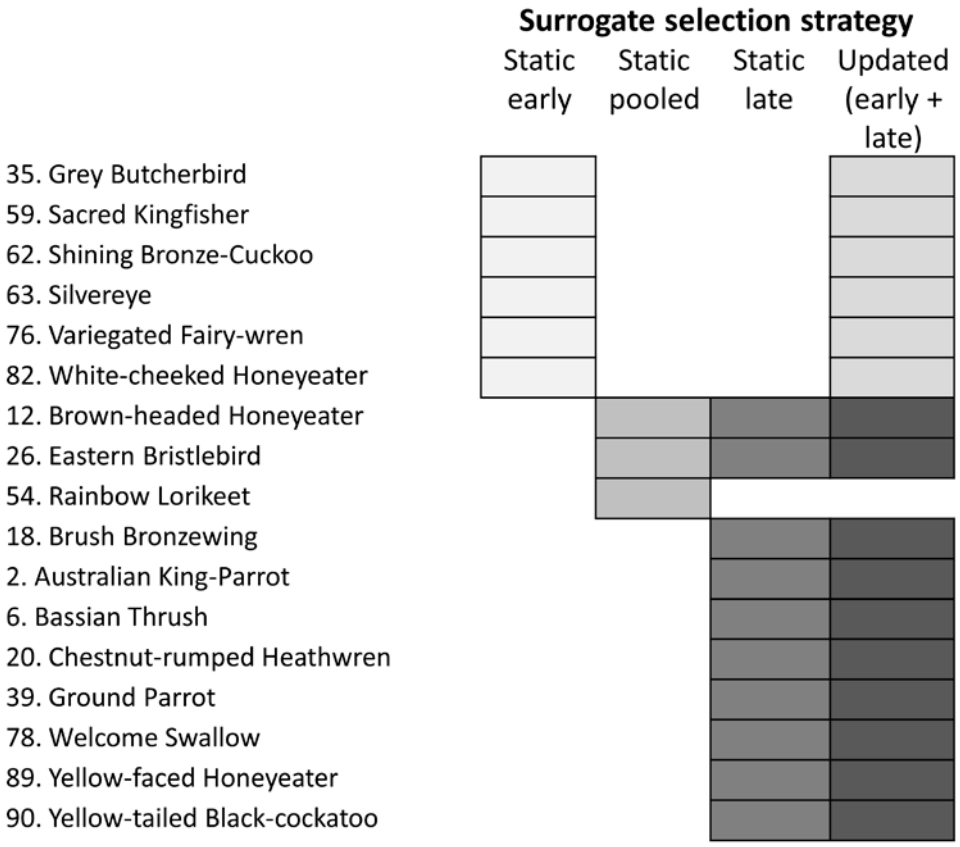


Figure 2. Composition of best surrogate sets for monitoring datasets of increasing temporal coverage in (a) the Southwest Slopes plantings and (c) Booderee National Park heathland, showing the best single surrogate (dark grey boxes) and the best complementary set of surrogates for achieving 70% surrogacy power of the entire community across all time (light grey boxes). (b and d) Dendrogram of Bray-Curtis dissimilarity indices comparing the species selected for achieving 70% surrogacy power in each scenario of co-occurrence dynamics resulting from community succession, season, or alternative allocations of sampling effort, in (b) the Southwest Slopes and (d) Booderee National Park.



705

706 **Figure 3.** Pareto-optimal curves trading off the number of selected surrogates against survey effort to
707 inform decisions under alternative monitoring protocols in (a) the Southwest Slopes (10 scenarios) and (b)
708 the Booderee National Park heathland (8 scenarios). Pareto frontiers are shown for achieving 70%
709 surrogacy power (dotted line, open data points) and 90% surrogacy power (solid line, closed data points)
710 across the entire succession (11 years). Each point represents the optimal decision from a given
711 monitoring scenario (1-3, 1-5, 1-7, 1-9 years (circles of increasing radius), 6-11 (small square) or 1-11
712 (large square) years, 1 in 2 years (triangle), or random (diamond) monitoring). Spring and winter
713 monitoring in the Southwest Slopes are shown by x and + respectively.



Surrogacy power				
Early period co-occurrence (1-5yrs)	70.2	56.5	69.0	**77.5**
Late period co-occurrence (6-11yrs)	51.1	54.9	69.6	**72.2**
All time co-occurrence (11yrs)	75.6	70.6	82.9	**90.3**

Figure 4. Comparison of performance of strategies for selecting surrogates that ignore co-occurrence dynamics (static strategies) versus account for dynamics (updated strategy). From left to right, the first strategy represents choosing a static set of complementary surrogate species representing 70% of early post-disturbance co-occurrences (monitoring species 1-5yrs after disturbance), and maintaining this surrogate set over the next 5 years. The second strategy represents waiting to select the optimal surrogates representing 70% of pooled co-occurrence across all time until 11 years of monitoring data are available. The third strategy represents choosing a static set of optimal surrogates representing 70% of late post-disturbance co-occurrences (using data collected 6-11 years post-disturbance and in ignorance of early co-occurrences). The final updated strategy chooses the optimal surrogates with 70% surrogacy power over the first 5 years then adds to this list the optimal surrogates with 70% surrogacy power over the next 5 years. By accounting for dynamic co-occurrence, this strategy always outperforms the other strategies in short-term and long-term surrogacy power (best-performing surrogate set across early, late and all time highlighted in bold).